

# Xuanjun Chen

✉ d12942018@ntu.edu.tw | 🏠 xjchen.tech | 🐙 xjchenGit | 📺 jun-ntu | 📄 Google Scholar

## Education

### National Taiwan University (NTU)

PH.D. IN COMMUNICATION ENGINEERING, EECS COLLEGE (GPA 4.30/4.30)

M.S. IN COMPUTER SCIENCE AND INFORMATION ENGINEERING (GPA: 4.19/4.30)

- **Advisors:** Prof. Jyh-Shing Roger Jang and Prof. Hung-yi Lee

Taipei Taiwan

Sep. 2023 - Present

Sep. 2020 - Jan. 2023

### National Taiwan University of Science and Technology (NTUST)

B.S. IN COMPUTER SCIENCE AND INFORMATION ENGINEERING (GPA: 4.11/4.30)

Taipei Taiwan

Sep. 2018 - Jun. 2020

### Shenzhen Univ. of Information Technology (f.k.a. Shenzhen Institute of Information Technology, SZIT)

JUNIOR COLLEGE IN ELECTRONIC INFORMATION ENGINEERING TECHNOLOGY

Shenzhen China

Sep. 2015 - Jul. 2018

## Honors

**First-Author Publications:** 1 × ICASSP, 2 × Interspeech, 1 × IEEE SLT, 1 × ISCA SPSC

**Scholarships:** Google Student Travel Grant ('24) | CTCI Bursary Award ('24) | 5 × Cantonese Scholarship ('20-'25) | 3 × Certificate of Achievement (Top 5%, '18-'20) | 3rd Prize Academic Award (Top 20%, '17) | National Encouragement Scholarship (Top 3%, '16)

**Competitions:** Ranked 3rd out of 42 teams in the logical access track of the ASVspoof 2021 challenge, Interspeech 2021 ('21) | National Bronze and Provincial Gold Awards, 3rd China College Internet Entrep. Comp. ('17)

## Experience

### Speech Processing and Machine Learning Lab, NTU

Sept. 2023 - Present

RESEARCH ASSISTANT, Advisor: Prof. Hung-yi Lee

- **[Research]** Researched on **Neural Audio Codec Model** and **Multimodal Large Language Model**
  - Co-built a large-scale spoken language model benchmark, ensuring standardized evaluation tasks. [ICLR '25]
  - Co-built a neural audio codec benchmark [ACL Findings '24] and co-organize a neural audio codec challenge [SLT '24].
  - Contributed to an audio language modeling report comparing methodologies systematically. [arXiv '24]
  - Contributed to a technical report on spoken language models for neural audio codec evaluation. [arXiv '24]
- **[Project]** Leading a student AI team in **TAIHUCAIS: TAIwan Humanities Conversational AI Knowledge Discovery System**
  - Developing a multi-agent system for humanities research scenarios based on Large Language Models [Demo].

### Multimedia Information Retrieval Lab, NTU

Oct. 2020 - Present

RESEARCH ASSISTANT, Advisor: Prof. Jyh-Shing Roger Jang

- **[Research]** Researched on **Multimodal Biometric Authentication**. Co-Advisor: Prof. Hung-yi Lee; Mentor: Dr. Haibin Wu
  - Proposed a effective & efficient audio-visual active speaker detection model via Transformer distillation [ICASSP 2024]
  - Proposed a SingGraph model using music and speech foundation models for SingFake detection [Interspeech 2024]
  - Proposed neural codec-based adversarial sample detection method for speaker verification [Interspeech 2024]
  - Built a diffusion and flow-matching based deepfake speech dataset [SLT 2024]
  - Analyzed model vulnerabilities and proposed a Push-Pull loss function against adversarial attack [SLT 2022]
  - Proposed an effective and efficient anti-spoofing model via two-stage distillation schema [ISCA SPSC 2022].
- **[Research]** Researched on **Speech Enhancement** and **Source Separation**
  - Proposed a MAMBA-based multichannel speech enhancement model using spectral-spatial information [ICASSP 2025].
  - Built a singer separation dataset for karaoke content generation [O-COCOSDA 2024].

## Service

2023-Now	<b>Reviewer / Program Committee</b> , EMNLP (2025), ACL (2024), ICASSP (2023-2025), LREC-COLING/COLING (2024-2025), MLSP (2024), IALP (2024), ECCV AVGenL (2024)
2023-Now	<b>Administrative Assistant</b> , NVIDIA-NTU Artificial Intelligence Joint Research Center, NTU
2024	<b>Invited Talker</b> , Topic: "Singing Voice Graph Modeling for SingFake Detection", Special Session: SVDD @ IEEE SLT 2024
2024	<b>Technical Committee</b> , Codec-SUPERB Challenge at IEEE 2024 Spoken Language Technology Workshop (SLT 2024)
2024-2025	<b>Teaching Assistant</b> , EE5200: <a href="#">Introduction to Generative AI (2024 Spr.)</a> and EE5184: <a href="#">Machine Learning (2025 Spr.)</a> , NTU

## Open Source and Skills

**Open Source** Official repository of MTDVocaLiST model for Audio-Visual Synchronization task (Starred 24)

**Open Source** Official repository of SingGraph model for SingFake Detection task (Starred 23, Fork 4)

**PLs and Toolkits** Python, LaTeX, Shell, PyTorch, Huggingface, PyTorch Lightning, Numpy, Pandas, Matplotlib.

**Language** Mandarin (native), Cantonese (native), English (fluent).

## Publications <sup>(† indicates equal contribution.)</sup>

---

- [1] **Xuanjun Chen**<sup>†</sup>, Jiawei Du<sup>†</sup>, Haibin Wu, Lin Zhang, I-Ming Lin, I-Hsiang Chiu, Wenze Ren, Yuan Tseng, Yu Tsao, Jyh-Shing Roger Jang, Hung-yi Lee, “[CodecFake-Omni: A Large-Scale Codec-based Deepfake Speech Dataset](#)” *Submitted to TASLP, 2025.*
- [2] Chien-yu Huang, Wei-Chih Chen, Shu-wen Yang, Andy T. Liu, Chen-An Li, Yu-Xiang Lin, Wei-Cheng Tseng, Anuj Diwan, Yi-Jen Shih, Jiatong Shi, William Chen, **Xuanjun Chen**, et al., “[Dynamic-SUPERB Phase-2: A Collaboratively Expanding Benchmark for Measuring the Capabilities of Spoken Language Models with 180 Tasks](#)”. in *International Conference on Learning Representations (ICLR), Singapore, April 2025.*
- [3] Chih-Kai Yang, Yu-Kuan Fu, Chen-An Li, Yi-Cheng Lin, Yu-Xiang Lin, Wei-Chih Chen, Ho Lam Chung, Chun-Yi Kuan, Wei-Ping Huang, Ke-Han Lu, Tzu-Quan Lin, Hsiu-Hsuan Wang, En-Pei Hu, Chan-Jan Hsu, Liang-Hsuan Tseng, I-Hsiang Chiu, Ulin Sanga, **Xuanjun Chen**, Po-chun Hsu, Shu-wen Yang, Hung-yi Lee, “[Building a Taiwanese Mandarin Spoken Language Model: A First Attempt.](#)” *Tech Report, Nov. 2024.*
- [4] Wenze Ren, Haibin Wu, Yi-Cheng Lin, **Xuanjun Chen**, Rong Chao, Kuo-Hsuan Hung, You-Jin Li, Wen-Yuan Ting, Hsin-Min Wang, Yu Tsao, “[Leveraging Joint Spectral and Spatial Learning with MAMBA for Multichannel Speech Enhancement.](#)” in the *2025 International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2025), Hyderabad, India, April 2025.*
- [5] **Xuanjun Chen**, Haibin Wu, Chung-Che Wang, Hung-yi Lee, Jyh-Shing Roger Jang “[Multimodal Transformer Distillation for Audio-Visual Synchronization.](#)” in the *2024 International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2024), Seoul, Korea, April 2024.*
- [6] **Xuanjun Chen**, Haibin Wu, Jyh-Shing Roger Jang, Hung-yi Lee “[Singing Voice Graph Modeling for SingFake Detection.](#)” in *Interspeech 2024, Kos Island, Greece, Sept 2024.*
- [7] **Xuanjun Chen**<sup>†</sup>, Jiawei Du<sup>†</sup>, Haibin Wu, Jyh-Shing Roger Jang, Hung-yi Lee “[Neural Codec-based Adversarial Sample Detection for Speaker Verification.](#)” in *Interspeech 2024, Kos Island, Greece, Sept 2024.*
- [8] Haibin Wu, **Xuanjun Chen**, Yi-Cheng Lin, Kai-wei Chang, Ho-Lam Chung, Alexander H. Liu, Hung-yi Lee “[Towards audio language modeling-an overview.](#)” *Overview Report, Feb. 2024.*
- [9] Haibin Wu, Ho-Lam Chung, Yi-Cheng Lin, Yuan-Kuei Wu, **Xuanjun Chen**, Yu-Chi Pai, Hsiu-Hsuan Wang, Kai-Wei Chang, Alexander H. Liu, Hung-yi Lee “[Codec-SUPERB: An In-Depth Analysis of Sound Codec Models.](#)” in *Findings of the Association for Computational Linguistics ACL 2024, Bangkok, Thailand, Aug 2024.*
- [10] Haibin Wu, **Xuanjun Chen**, Yi-Cheng Lin, Jiawei Du, Kai-Wei Chang, Ke-Han Lu, Alexander Liu, Ho-Lam Chung, Yuan-Kuei Wu, Dongchao Yang, Songxiang Liu, Yi-Chiao Wu, Xu Tan, James Glass, Shinji Watanabe, and Hung-yi Lee “[Codec-SUPERB @ SLT 2024: A lightweight benchmark for neural codec models.](#)” in the *2024 IEEE Spoken Language Technology Workshop (IEEE SLT 2024), Macao, China, 2024.*
- [11] Jiawei Du, I-Ming Lin, I-Hsiang Chiu, **Xuanjun Chen**, Haibin Wu, Wenze Ren, Yu Tsao, Hung-yi Lee, Jyh-Shing Roger Jang “[DFADD: The Diffusion and Flow-Matching based Audio Deepfake Dataset.](#)” in the *2024 IEEE Spoken Language Technology Workshop (IEEE SLT 2024), Macao, China, 2024.*
- [12] Hsuan-Yu Lin, **Xuanjun Chen**, Jyh-Shing Roger Jang “[Singer separation for karaoke content generation.](#)” in *The 27th Oriental COCOSDA Conference (O-COCOSDA 2024), Hsinchu, Taiwan, Oct 2024.*
- [13] **Xuanjun Chen**<sup>†</sup>, Haibin Wu<sup>†</sup>, Helen Meng, Hung-yi Lee, Jyh-Shing Roger Jang “[Push-Pull: Characterizing the Adversarial Robustness for Audio-Visual Active Speaker Detection.](#)” in the *2022 IEEE Spoken Language Technology Workshop (IEEE SLT 2022), Doha, Qatar, Jan 2023.*
- [14] **Xuanjun Chen**<sup>†</sup>, Yen-Lun Liao<sup>†</sup>, Chung-Che Wang, Jyh-Shing Roger Jang “[Adversarial Speaker Distillation for Counter-measure Model on Automatic Speaker Verification.](#)” in *Proc. 2nd Symposium on Security and Privacy in Speech Communication (SPSC 2022), Incheon, Korea, Sept 2022.*